

Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.



United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

NASS Research Report
Number SRB-88-04

May 1988

Sto

A Test for Bias in the Location of Sunflower OY Units

Antoinette Tremblay
Ron Fecso

A TEST FOR BIAS IN THE LOCATION OF SUNFLOWER OY UNITS by
Antoinette Tremblay and Ron Fecso, Research and Applications
Division, National Agricultural Statistics Service, U.S.
Department of Agriculture, Washington, D.C. 20250, May 1988.
NASS Research Report No. SRB-88-04.

ABSTRACT

Randomness of the objective yield unit location for sunflower is examined in this nonsampling error study. A test is developed based on the hypothesis that the starting and ending points of the unit should be uniformly distributed between their adjacent sunflower plants. Goodness-of-fit tests on 1984 data reject the hypothesis that placement of the starting stake of the unit is following a random process while accepting the ending stake's placement as basically following a random process. These results raise concern that the unit is not being randomly located within the row. This paper also presents additional information regarding the distribution of spacings between sunflower plants and the distribution of lengths of the count unit. Major recommendations as a result of this study include increased attention to unit location at survey schools and periodic quality control studies to examine bias.

KEY WORDS

chi-square goodness-of-fit test, objective yield survey, quality control, unit location bias

* This paper was prepared for limited distribution to *
* the research community outside the U.S. Department of *
* Agriculture. The views expressed herein are not *
* necessarily those of NASS or USDA. *

ACKNOWLEDGMENTS

The authors thank the Minnesota, North Dakota and South Dakota State Statistical Offices for their efforts in the extra data collection required to conduct this research study. We also thank Doug Bond and Sally Amedo for their initial analyses, and Read Johnston for solving computer problems along the way.

Washington, D.C.

May, 1988

CONTENTS

Page

| | |
|--|-----|
| SUMMARY | iii |
| INTRODUCTION..... | 1 |
| PROCEDURES..... | 2 |
| Objective Yield Sample..... | 2 |
| Within Field Procedures..... | 3 |
| Research Procedures..... | 3 |
| FORMULATION OF THE TEST STATISTIC..... | 6 |
| RESULTS..... | 8 |
| EXPLORATORY ANALYSIS..... | 9 |
| Between-Plant Distances..... | 9 |
| Length of Fifteen-Foot Count Unit..... | 12 |
| CONCLUSIONS AND RECOMMENDATIONS..... | 14 |
| REFERENCES..... | 16 |
| APPENDIX A: OY SURVEY AND RESEARCH MEASUREMENTS FORMS..... | 17 |
| APPENDIX B: ILLUSTRATIONS OF GOODNESS-OF-FIT TESTS..... | 20 |
| APPENDIX C: THEORETICAL VS. OBSERVED STAKE PLACEMENTS..... | 23 |

SUMMARY

This nonsampling error study on the 1984 sunflower objective yield (OY) survey examines one part of the sampling process--unit location. Procedures for unit location can be interpreted statistically as a hypothesis that placement of an OY unit's starting stake follows a uniform distribution along the row. This implies that all possible locations of the starting stake are equally likely with respect to the first plant in the unit. If these conditions are not met, a bias in the inclusion or exclusion of plants could result, possibly introducing bias in the yield estimate. To test the hypothesis on an initial sample of 205 sunflower OY survey sample fields, a chi-square goodness-of-fit test is developed for the distribution of field measurements.

The test rejects the hypothesis that placement of the starting stake of the unit is following a random process. Thus, the unit is not being randomly located within the row. Causes may include inadequate training or supervision, misunderstanding of or nonadherence to instructions, measurement difficulties in tape usage, rounding stalk width, or a combination of all. The placement of an OY unit's ending stake was also analyzed, revealing that it is following a random process. This result questions the effectiveness of the 5-foot buffer procedure. Location after the 5-foot buffer is not random, yet location 15 feet further, i.e., a 20-foot buffer, is random.

While undertaking the analyses, interesting statistical information regarding the distribution of spacings between sunflower plants and the distribution of lengths of the count unit became apparent. These topics are addressed briefly, while it is recommended that they be analyzed further.

Other recommendations as a result of this study include increased attention to these aspects of the survey process at training schools and in the field, periodic measurement of this process in sunflower and other crops, further analysis to attempt to measure plant inclusion/exclusion bias, and studies of alternate sampling methods which utilize the plant spacing distribution.

A TEST FOR BIAS IN THE LOCATION OF SUNFLOWER OY UNITS

Antoinette Tremblay and Ron Fecso¹

INTRODUCTION

"Objective yield surveys provide information for making estimates or forecasts of crop yield based directly on plant counts, measurements, and weights of the crop obtained from small plots in a random sample of fields" [7]². The plot (i.e., unit) data, collected in the final stage of a multi-stage sampling design, receive a large expansion in arriving at an estimate of gross yield per acre since the units are quite small in area. Any bias associated with the final stage has a disproportionate potential for influencing the validity of the survey data.

In 1984, the oilseed sunflower³ objective yield (OY) estimate was the subject of much concern, with a general perception that it was seriously biased when compared to the estimates published by the Agricultural Statistics Board of the National Agricultural Statistics Service (NASS), U.S. Department of Agriculture [1]. A quality control program for sunflower was recommended in order to ensure that OY survey procedures, e.g., unit location, are carefully followed. The bias was less severe in 1986, the result of changes to the estimating procedure. Yield is now determined by expanding seed weight per unit area rather than head weight per unit area. Still, studies of potential bias-inducing processes for all OY crops are worthwhile as means of quality control.

Zarkovich [9] discusses the concept of unit location as a source of bias, stating that "small plots might lead to biases because there is a kind of...psychological bias on the part of the investigator to include unduly some of the bordering plants or tillers inside a cut". He observed this same basic tendency of obtaining overestimation with a large number of surveys dealing with different crops. Generally, Zarkovich studied the biases of small units in broadcast fields. A similar bias problem may occur if the units laid out in row-planted fields are not located at random. This nonsampling error study analyzes sunflower OY data, but the placement problems may exist for many of the OY crops.

¹ The authors are mathematical statisticians with the National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C.

² Bracketed numbers refer to literature cited at the end of this report.

³ Hereafter referred to as sunflower.

NASS OY procedures provide for a random-row, random-pace within row selection method. Plot-size biases or row location are not addressed herein, only the placement of the unit within the row. An important consideration, therefore, is that the starting point of the unit be randomly located in the row. The testing of the randomness hypothesis involves the implementation of chi-square goodness-of-fit tests to determine whether the enumerator lays out the unit according to prescribed instructions. This paper also tests the randomness hypothesis as applied to placement of the unit's ending stake. Biases may exist if either hypothesis cannot be accepted, resulting in over- or under-estimation.

While undertaking any analysis, interesting and additional statistical information often becomes apparent. In this study, the distribution of space between sunflower plants, and the distribution of lengths of the fifteen-foot count unit are two such areas. They are discussed in "Exploratory Analysis".

PROCEDURES

Objective Yield Sample

In 1984, 205 sunflower OY survey sample fields were subsampled from the June Enumerative Survey (JES). Tracts with sunflower acreage in the producing states of Minnesota, North Dakota and South Dakota composed the sampling frame⁴. Samples were selected with probability proportional to size (PPS), based on expanded tract acreage. Table 1 specifies the resultant sample size of 165 fields. Individual state analyses are not attempted in this study.

TABLE 1: Samples for the 1984 Sunflower OY Survey

| Sample Status | States | | | Total |
|--------------------------|----------------------|-------------|------------|-------------|
| | MINN | N DAK | S DAK | |
| Farmer refusal | 2 (8.0) ⁵ | 5 (4.0) | 0 (0.0) | 7 (3.4) |
| No sunflower in tract | 0 (0.0) | 5 (4.0) | 9 (16.4) | 14 (6.8) |
| Sunfl'r not for oilseed | 8 (32.0) | 5 (4.0) | 0 (0.0) | 13 (6.4) |
| Lost sample | 0 (0.0) | 1 (0.8) | 0 (0.0) | 1 (0.5) |
| No research measurements | 0 (0.0) | 0 (0.0) | 5 (9.1) | 5 (2.4) |
| Row-planted sunflower | 15 (60.0) | 109 (87.2) | 41 (74.5) | 165 (80.5) |
| TOTAL | 25 (100.0) | 125 (100.0) | 55 (100.0) | 205 (100.0) |

⁴ 1986 was the first year in which sunflower OY sample fields were subsampled from a combination of the list frame and JES area frame samples, i.e., the multiple frame survey.

⁵ Numbers in parentheses are percentages of the state total.

Within Field Procedures

Details and procedures of the operational sunflower OY program are documented in Section 4 of the Enumerator's Manual [8], with the steps relevant to this study restated here.

At the time of the initial interview (approximately August 1, i.e., 8/1) and after having received permission to do so, the enumerator independently locates and lays out two units in each sunflower sample field. In locating unit 1 from the selected point of field entry, the enumerator walks the preassigned random number of rows along the field, ending at row 1 of unit 1; the next row in the direction of travel is row 2 of unit 1. The enumerator then walks the preassigned random number of paces into the field along the middle between rows 1 and 2. After the last pace, the enumerator places a dowel stick across rows at right angles to the direction of the rows. The zero end of a 50-foot steel tape is positioned immediately beyond the dowel stick so that the tape may be drawn beside the plants in row 1, with an anchor stake placed at this zero end. (Figure 1 illustrates these OY procedures.) A starting stake for row 1 of unit 1 is positioned exactly 5 feet from the anchor stake, creating a 5-foot buffer which is intended to be a control procedure to decrease the potential for inadvertent subjectivity in the unit location procedures by the enumerator. An ending stake is then placed exactly at the 20-foot mark. Hereafter, these starting and ending stakes are called the 5-foot and 20-foot stakes, respectively. For row 2 of unit 1, the enumerator anchors the 50-foot tape beyond the dowel stick beside the plants in row 2. Five-foot starting and 20-foot ending stakes are positioned as such.

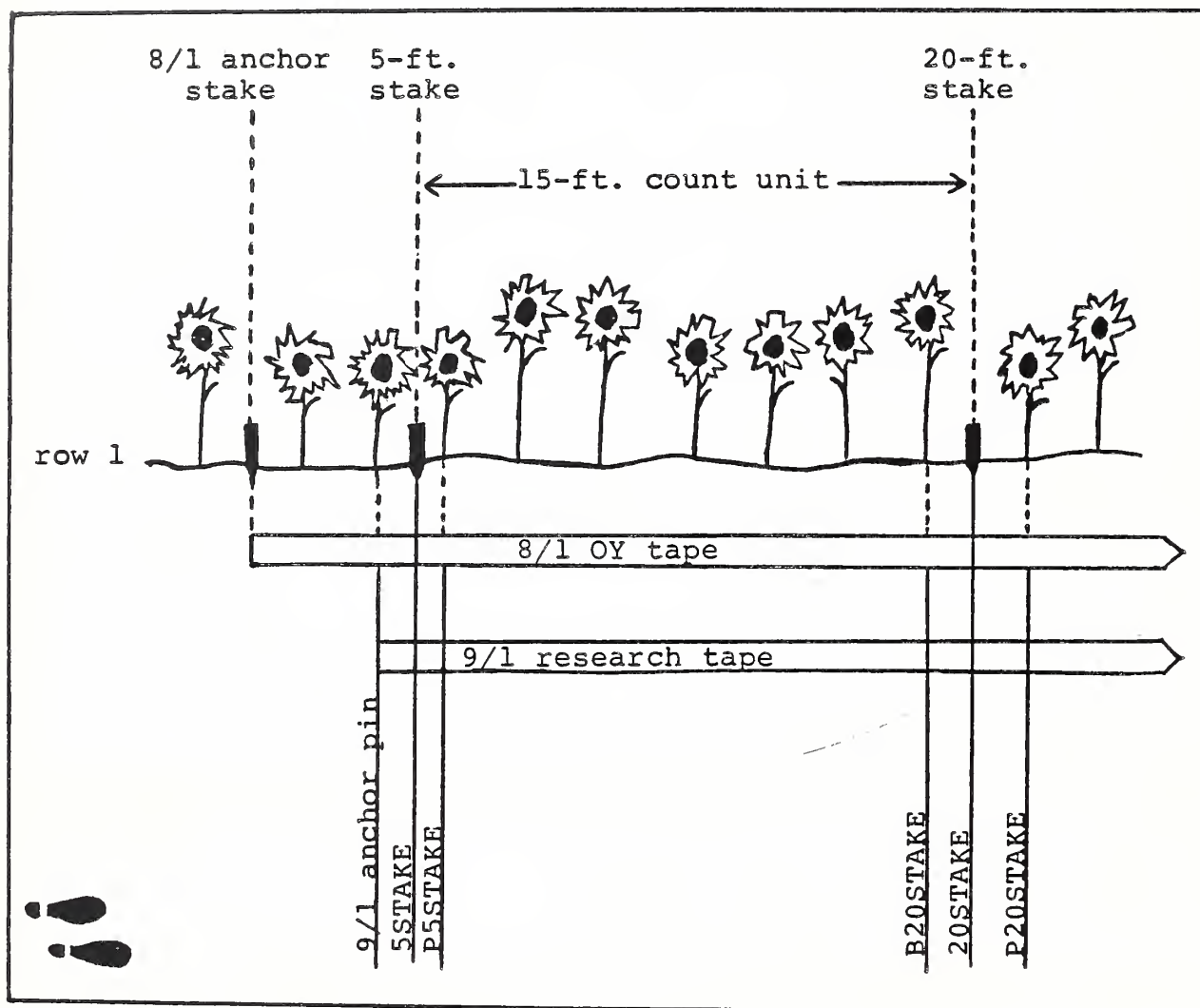
The 5- and 20-foot stakes of each row define the 15-foot x 2-row count unit. Unit 1 counts and measurements are recorded on the Form B (see Appendix A). Unit 2 is likewise located in relation to the starting corner, and its associated data recorded on the same Form B.

Research Procedures

The procedures for the research measurements required by Form I: Sunflower Research Measurements, 1984 (see Appendix A) are documented in Section 6 of the Enumerator's Manual [8] and are also restated here.

During the September 1, 1984 visit only, and before beginning the count of stalks as required by item 3 of Form B, the enumerator was to record the measurements relevant to placement

FIGURE 1: OY and Research Measurements⁶,
1984 Sunflower Research



⁶ The 9/1 research tape, 9/1 anchor pin, and variables 5STAKE, P5STAKE, B20STAKE, 20STAKE and P20STAKE are described in the section "Research Procedures". Also, a stake which divides row 1 into clip areas A and B is omitted from the figure since it has no relevance to the scope of this study.

of the 5- and 20-foot stakes. Measurements on 660 rows were to be made -- 165 sunflower sample fields (see Table 1) x 2 units/sample field x 2 rows/unit. These measured rows are called observations⁷. Figure 1 illustrates the research measurements and the corresponding variable names which are described below⁸.

- (1) "Anchor the zero end of the 50 ft. steel tape exactly at the last sunflower stalk (dead or alive) before the 5 ft. stake. [This anchor is called the 9/1 anchor pin.] Stretch the steel tape past the first stalk which is just past the 20 ft. florist stake."
- (2) "Across the columns of Form I, record measurements to the nearest 0.1 ft. at the following points: [The corresponding variable names used in this paper's analysis are in brackets.]
 - a. The 5 ft. stake [5STAKE].
 - b. The first stalk past the 5 ft. stake [which is the first stalk of the unit for the specified row [P5STAKE]].
 - c. The last stalk before the 20 ft. stake [which is the last stalk of the unit for the specified row [B20STAKE]].
 - d. The 20 ft. stake [20STAKE].
 - e. The first stalk past the 20 ft. stake [P20STAKE]."

- (3) Continue with item 3 of Form B.

Section 6 of the Enumerator's Manual also provides the enumerator with procedures to undertake when measurement problems arise. For example, procedures to follow when there are no stalks between the 5- and 20-foot stakes or when a stalk lines up exactly with a florist stake. Measurements that resulted from use of these special procedures were examined in the analysis and found to have no impact on the results.

⁷ To test the hypothesis, 655 observations are usable since they have nonmissing and consistent measurements. An additional five observations having unusable measurements are deleted when placement of the 20-foot stake is analyzed.

⁸ Except for words in brackets [], quoted material is from Section 6 of the Enumerator's Manual.

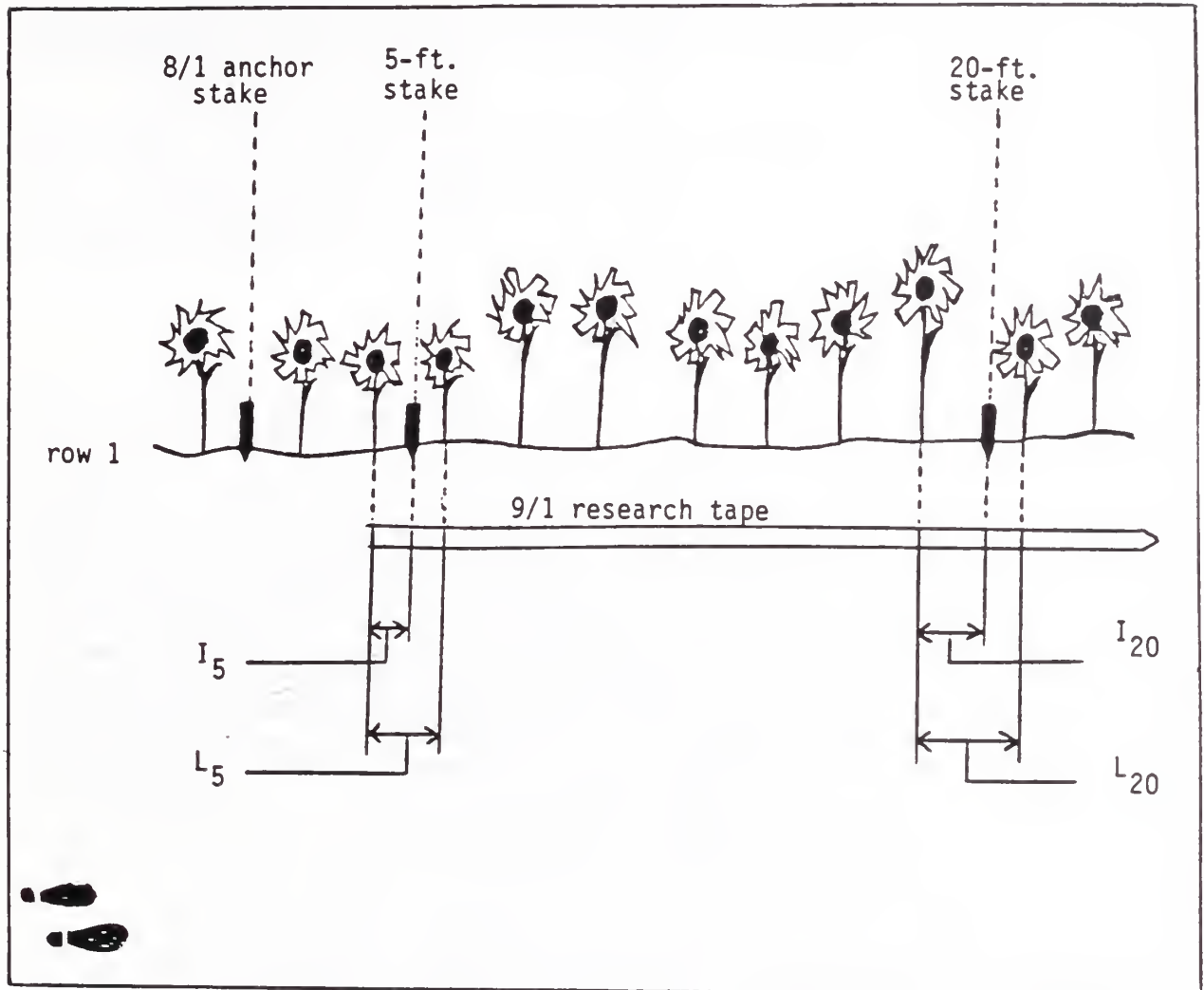
FORMULATION OF THE TEST STATISTIC

If survey procedures are followed, it is hypothesized that placement of an OY unit's starting stake follows a uniform distribution along the row. This being true, then there should not exist any bias on the part of the enumerator to place a stake closer or further from the plants on either side of the stake. Referring to Figure 2, this states that the 5-foot stake should have an equal probability of being placed anywhere in L_5 . If this condition is not met, the 15-foot length of the unit could vary and a bias in the inclusion/exclusion of plants could result.

Chi-square goodness-of-fit tests are usually used in this situation to test whether empirical data are compatible with a specified theoretical distribution [4]. The chi-square test statistic is based on the assumption of independence of the trials, an incorrect assumption since adjacent rows are obviously dependent. However, it is assumed that the two units in the field and the systematically selected fields are essentially independent repetitions of the process of placing the stake in a row. Therefore, the main test will be made using the 5-foot stake measurements from all row 1 observations treated as independent selections. Row 2 and 20-foot stake analyses are made and presented to help explain location procedure problems. An adequate simultaneous testing procedure is not immediately available and the resources needed to develop such a test were deemed excessive since the main hypothesis will be shown to be rejected.

The chi-square test developed is not a straightforward test for a uniform distribution because the field measurements come from plant spacings which are variable and the measurements of these distances are rounded. Over all observations, the interval L_5 ranges from 0.1 to 18.5 feet with a mean of 1.9 feet; the interval L_{20} ranges from 0.1 to 21.6 feet with a mean also of 1.9 feet. These varying measurements may be the result of problems with the use of the tape, misunderstanding of or nonadherence to instructions, rounding of the stalk width, or a combination of all. Also, and very important to the analysis, the measurements of these distances are rounded since the 50-foot steel tape is scaled in increments of 0.1 feet. Thus, the theoretical distribution of measurements is discrete relative to the L_5 and L_{20} lengths, and is not close enough to an exact uniform distribution for a straightforward use of chi-square distribution tests. Therefore, the ratio of stake placement to the distance between the stake's adjacent plants will provide the basis for testing the assumption of a uniform distribution. The hypothesis translates into one of testing if I_5/L_5 and I_{20}/L_{20} are uniformly distributed among ten subintervals in the

FIGURE 2: Stake Placement Ratio Variables⁹,
1984 Sunflower Research



⁹ These stake placement ratio variables can be defined in terms of the research measurements (see Figure 1): $I_5 = 5\text{STAKE}$, $L_5 = P5\text{STAKE}$, $I_{20} = 20\text{STAKE} - B20\text{STAKE}$, $L_{20} = P20\text{STAKE} - B20\text{STAKE}$.

[0,1] interval. They are:

[0,.1), [.1,.2), [.2,.3), ..., [.8,.9), [.9,1.0].

This paper also tests if I_5/L_5 and I_{20}/L_{20} are uniformly distributed among five subintervals in the [0,1] interval. This is done to assess the possibility of a very specific bias--placing the stake in the middle of the distance between the adjacent plants.

For the reader who wishes more specifics to the exact test, two illustrations in Appendix B present how the discrete measurements affect the uniform distribution assumption, thus requiring the creation of a distribution which reflects the interaction between rounding and the variable plant spacing.

RESULTS

Table 2 provides the p-values¹⁰ for the eight individual chi-square goodness-of-fit tests of the assumed distribution. The major hypothesis test, the 5-foot stake placement ratio, is significantly different than the theoretical unit interval distribution for both rows of both subinterval cases. In fact, placement of the 5-foot stake is biased away from the last plant before the unit. (Appendix C Figures 1-16 examine this in more detail; they provide comparisons of the expected (theoretical) and observed stake placements, by 5- and 20-foot stake, by unit and row, and by 10 and 5 subintervals.) Thus, it appears that a nonrandom process is being practiced with the specified p-values [3]. This nonrandomness of the 5-foot stake placement may be due to improper unit layout, for a variety of reasons such as nonadherence to or misunderstanding of instructions, measurement difficulties with tape usage, buffer effects, ... The effects of this bias in terms of plant inclusion/exclusion will have to come from a subsequent research study which will merge this research data with the accompanying OY plant count data.

Since the results are similar for both subinterval cases, rounding in the measurement process is not a cause for the significance.

The 20-foot stake placement ratio is, basically, not significantly different than the theoretical unit interval distribution. (The low p-value for 10 subintervals, unit 1, row 1, cannot be explained.)

¹⁰ A p-value is the probability under the null hypothesis of obtaining a result equal to, or more extreme than, the observed.

TABLE 2: Chi-Square Components for the 5-Foot and 20-Foot Stake Data Sets, 1984 Sunflower Research

| Data set | -----Pr>X ² ----- | | |
|----------------------|------------------------------|--------|------------|
| | unit 1 | unit 2 | both units |
| 10 subintervals | | | |
| 5-foot stake: row 1 | .035 | .166 | .006 |
| row 2 | .011 | .081 | .041 |
| 20-foot stake: row 1 | .081 | .369 | .025 |
| row 2 | .440 | .254 | .254 |
| 5 subintervals | | | |
| 5-foot stake: row 1 | .013 | .365 | .010 |
| row 2 | .126 | .041 | .051 |
| 20-foot stake: row 1 | .919 | .488 | .464 |
| row 2 | .442 | .435 | .521 |

As mentioned, these results question the effectiveness of the 5-foot buffer procedure, which is implemented to reduce unit location bias. This sunflower unit location bias study shows that the location after the 5-foot buffer may not be random, yet the location 15 feet further, which could be thought of as a 20-foot buffer, is random. Also, in a wheat OY study on buffer zone effects, Matthews [2] observed a trend towards fewer heads counted in buffered sample units, although the differences were not significant. Therefore, it is questioned if the purpose of the 5-foot buffer is being fulfilled.

EXPLORATORY ANALYSIS

Although a research project needs to define the tests associated with to-be-collected data to ensure proper inference, it is usually the case that interesting statistical information becomes apparent during the analysis of the data. This section highlights some of this exploratory data and proposes additional analysis for which the data might be useful.

Between-Plant Distances

An interesting observance is the distribution of space between plants in a row (Figures 3 and 4). The shape of the distributions strongly suggests a distribution from the exponential family, e.g., lognormal. Similar distributions are used to develop optimal forest sampling plans, thus raising the question of whether similar plans would be useful in row crops.

FIGURE 3: Distribution of Space Between Sunflower Plants Adjacent to Five-Foot Stake (L₅),
1984 Sunflower Research

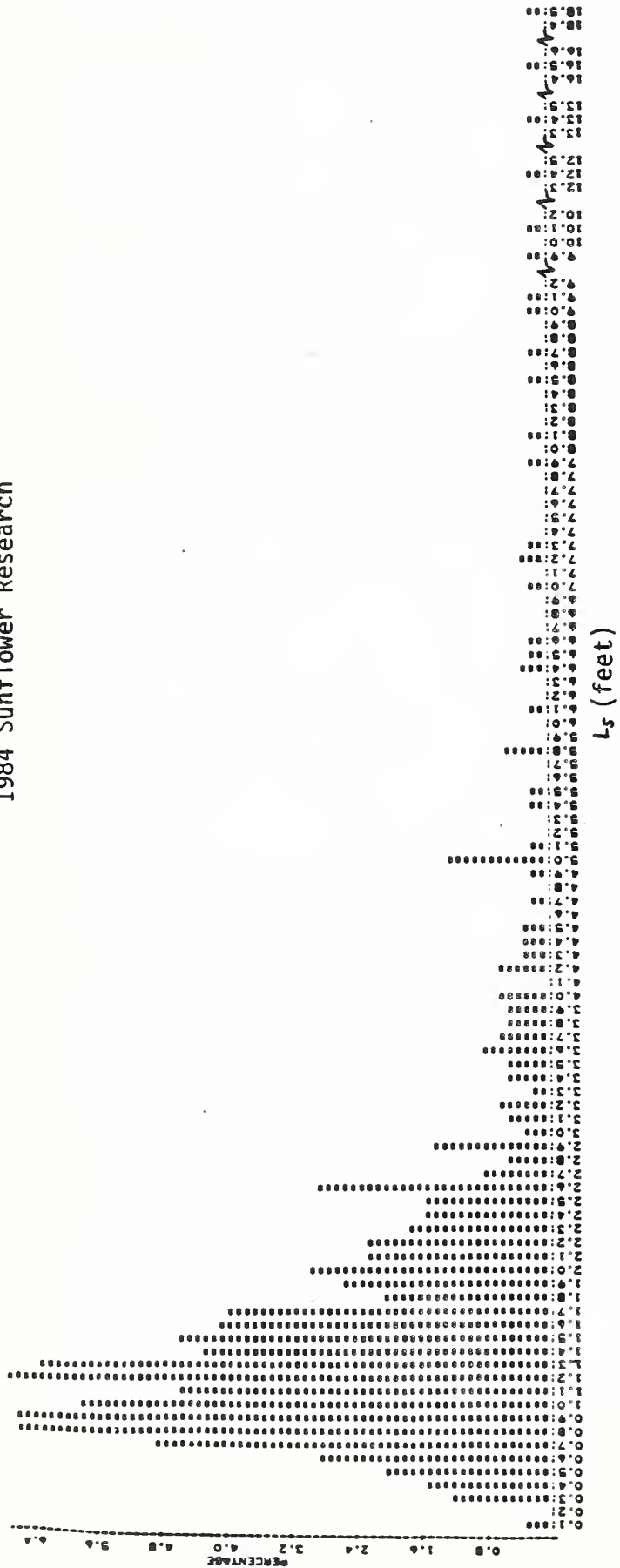
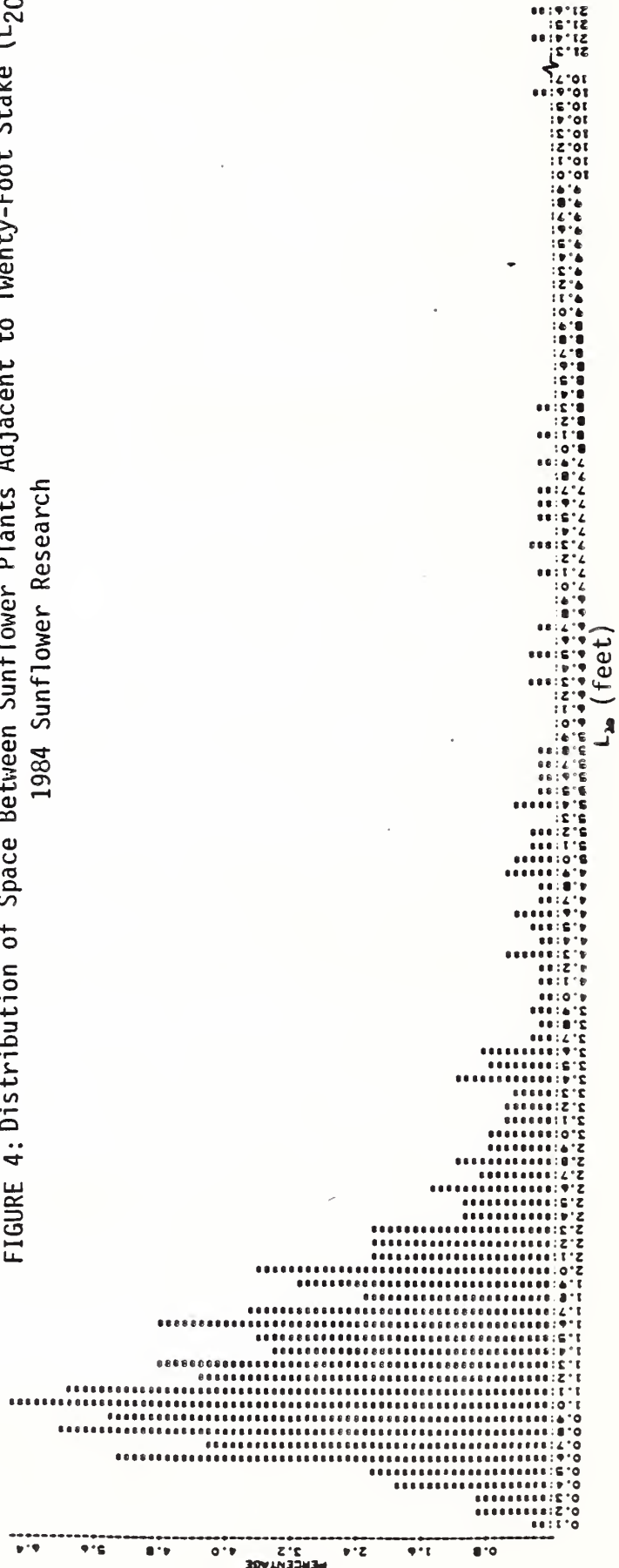


FIGURE 4: Distribution of Space Between Sunflower Plants Adjacent to Twenty-Foot Stake (L₂₀),
1984 Sunflower Research



Sopher and Ordokhani [5] found that the exponential distribution, with the mean between-plant distance as the exponential parameter (mean and standard deviation), fit the spacing of cotton plants very well and provided a very useful stand indicator for use by the farm operator. This type of information could lead to the development of improved sampling techniques, and feedback information to help gain farmer support.

The correlation matrix of between-plant distances (Table 3) supports the high variability.

Some enumerator effects were observed during the data editing. A reexamination of this data, testing for enumerator effects, could provide insight into the causes of the biases detected. For example, testing for enumerator specific differences between plant spacing at the beginning versus the end of the unit.

TABLE 3: Correlation Matrix of Between-Plant Distances, by stake, unit, and row, 1984 Sunflower Research

| | U1R15 | U1R25 | U2R15 | U2R25 | U1R120 | U1R220 | U2R120 | U2R220 |
|---------------------|-------|--------------------|-------|-------|--------|--------|--------|--------|
| U1R15 ¹¹ | 1.00 | .30* ¹² | .27* | .34* | .23 | .13 | .32* | .07 |
| U1R25 | .30* | 1.00 | .07 | .13 | .11 | .23 | .14 | .02 |
| U2R15 | .27* | .07 | 1.00 | .35* | .35* | .06 | .23 | .18 |
| U2R25 | .34* | .13 | .35* | 1.00 | .17 | .05 | .35* | .28* |
| U1R120 | .23 | .11 | .35* | .17 | 1.00 | .19 | .06 | .11 |
| U1R220 | .13 | .23 | .06 | .05 | .19 | 1.00 | .19 | .18 |
| U2R120 | .32* | .14 | .23 | .35* | .06 | .19 | 1.00 | .30* |
| U2R220 | .07 | .02 | .18 | .28* | .11 | .18 | .30* | 1.00 |

¹¹ U1R15 indicates 5-foot stake placement in row 1 of unit 1. Notation is similar for remaining components.

¹² An asterisk indicates significance at $\alpha=.001$.

Length of Fifteen-Foot Count Unit

The length of the 15-foot count unit is important to examine since the calculation of yield per acre assumes an exact 15-foot row, and deviations from 15 feet may result in a bias. When analyzing all 660 observations, the distance between the 5- and 20-foot stakes, as calculated from the Form I research measurements, ranges from 12.00 to 20.20 feet with a mean of 15.08, standard error of the mean of 0.03, and coefficient of variation (CV) of 4.74 percent. Figure 5 exhibits this distribution. The tails of the distribution point to some 6 percent of the measurements which have serious measurement discrepancies (greater than ± 0.5 feet), with this percentage being basically the same for both rows 1 and 2. The unit could possibly have been laid out correctly in relation to the adjacent plants on the first field visit and simply mismeasured on the research visit. Also, much of the deviation near 15 feet may be due to measurement errors caused by slackness in the tape, rounding, pulling the Sept. anchor pin slightly out, or interference from weeds or other field environmental conditions. In any event, when considering the number of non-15-foot rows indicated in Figure 5 and the generally small between-plant spacing seen in Figures 3 and 4, the possibility of plant inclusion/exclusion arises.

FIGURE 5: Distribution of 15-Foot Count Unit Distance Between
Stakes, 1984 Sunflower Research

| FIFTEEN | | FREQ | CUM FREQ | PCT | CUM PCT |
|---------|-------|------|-------------|-------|------------|
| . | | 6 | 6 | 0.91 | 0.91 |
| 12.0 | | 1 | 7 | 0.15 | 1.06 |
| 13.5 | | 1 | 8 | 0.15 | 1.21 |
| 13.6 | | 1 | 9 | 0.15 | 1.36 |
| 13.8 | | 1 | 10 | 0.15 | 1.52 |
| 14.0 | | 2 | 12 | 0.30 | 1.82 |
| 14.1 | | 2 | 14 | 0.30 | 2.12 |
| 14.2 | | 1 | 15 | 0.15 | 2.27 |
| 14.3 | | 4 | 19 | 0.61 | 2.88 |
| 14.4 | | 5 | 24 | 0.76 | 3.64 |
| 14.5 | * | 8 | 32 | 1.21 | 4.85 |
| 14.6 | | 6 | 38 | 0.91 | 5.76 |
| 14.7 | * | 14 | 52 | 2.12 | 7.88 |
| 14.8 | ** | 28 | 80 | 4.24 | 12.12 |
| 14.9 | ***** | 68 | 148 | 10.30 | 22.42 |
| 15.0 | ***** | 325 | 473 | 49.24 | 71.67 |
| 15.1 | ***** | 95 | 568 | 14.39 | 86.06 |
| 15.2 | *** | 43 | 611 | 6.52 | 92.58 |
| 15.3 | * | 16 | 627 | 2.42 | 95.00 |
| 15.4 | * | 10 | 637 | 1.52 | 96.52 |
| 15.5 | | 5 | 642 | 0.76 | 97.27 |
| 15.6 | | 3 | 645 | 0.45 | 97.73 |
| 15.7 | | 1 | 646 | 0.15 | 97.88 |
| 15.8 | | 1 | 647 | 0.15 | 98.03 |
| 17.6 | | 1 | 648 | 0.15 | 98.18 |
| 19.3 | | 1 | 649 | 0.15 | 98.33 |
| 19.8 | | 1 | 650 | 0.15 | 98.48 |
| 19.9 | | 2 | 652 | 0.30 | 98.79 |
| 20.0 | | 5 | 657 | 0.76 | 99.55 |
| 20.1 | | 2 | 659 | 0.30 | 99.85 |
| 20.2 | | 1 | 660 | 0.15 | 100.00 |

-----+-----+-----+-----+-----+

10 20 30 40 50

PERCENTAGE

CONCLUSIONS AND RECOMMENDATIONS

This nonsampling error study of the 1984 sunflower crop dealt with OY unit location. Since placement of the starting stake of the unit was not following a random process, it is questioned if the unit is being laid out according to prescribed instructions. Causes may include inadequate training, misunderstanding of or nonadherence to instructions, supervision for a "one shot" research study, measurement difficulties with tape usage, rounding stalk width, or a combination of all. Also, since the ending stake's placement is basically following a random process, these results question the effectiveness of the 5-foot buffer procedure.

It is unknown whether these errors are only measurement errors which merely vary between the adjacent plants, or if they result in mismeasured, bias-inducing units by including/excluding any 'non-unit' plants. In any event, evidence of a possible biasing condition exists.

Based on this study's results, the following recommendations are made:

1. Since the results do show evidence of unit location problems, the importance of following location procedures needs to be stressed at training schools, improvements on these testing methods should be attempted, and periodic quality control analyses of unit location should be carried out.
2. Perform analyses to determine if there is a bias in the number of plants included in the unit. That is, determine how the placement of the 5- and 20-foot stakes and the length of the count unit between the stakes affect plant count and, thus, yield. An initial evaluation of the potential for this work can be done by merging operational data with the research data set and continuing the exploratory analysis.
3. Consider the possibility of these unit location problems on crops other than sunflower. For example, similar problems may exist in the corn OY program. Also, it is suggested that a greater share of the enumerator's time be spent on unit location, which is critical to accurate plant population estimates. Perhaps some of the extra time spent on accurately locating an OY unit could be made up by eliminating any fruit counts or measurements that have been shown to have little yield forecasting power.

4. As introduced in "Exploratory Analysis", continue research regarding between-plant spacing, which may lead to improved and more efficient sampling and/or estimation techniques. Information on plant spacing could prove to be of major interest and value to farm operators. This work can begin by comparing the number of plants counted in the operational unit with the number that would be expected in the unit based on the between-plant distance. This would also shed light on the potential bias caused by rows which are significantly different than 15 feet in length.

REFERENCES

1. Bond, Douglas C. 1983 Sunflower Objective Yield Survey Research. U.S. Department of Agriculture, Statistical Reporting Service. 1983.
2. Matthews, Ralph V. Head Count Differences Between Sample Plots in the 1983-1984 Wheat Objective Yield Surveys. U.S. Department of Agriculture, National Agricultural Statistics Service. 1986.
3. Miller, Rupert G., Jr. Beyond ANOVA, Basics of Applied Statistics. John Wiley & Sons, Inc. 1986. p. 2.
4. Snedecor, G. W. and W.G. Cochran. Statistical Methods, 6th ed. Ames, Iowa: Iowa State University Press. 1971. pp. 236-238.
5. Sopher C. D. and Nasser Ordokhani. Use of the Exponential Distribution in Predicting Cotton Plant Stands. Proceedings of 1970 Beltwide Cotton Production Research Conferences.
6. U.S. Department of Agriculture, Statistical Reporting Service. Objective Yield Supervising and Editing Manual. 1984.
7. U.S. Department of Agriculture, Statistical Reporting Service. Scope and Methods of the Statistical Reporting Service. 1983.
8. U.S. Department of Agriculture, Statistical Reporting Service. Sunflower Objective Yield Survey - Enumerator's Manual. 1984.
9. Zarkovich, S. S. Quality of Statistical Data. Rome: FAO. 1966.

UNITED STATES DEPARTMENT OF AGRICULTURE
STATISTICAL REPORTING SERVICEO.M.B. Number 0535-0088
Expiration Date 7/31/86
C.E. 12-0035B**FORM B—SUNFLOWER YIELD SURVEY—1984**
COUNTS**MONTH CODE**August 11
September 1 ..2
October 13
November 1 ..4
December 1 ..5YEAR, CROP, FORM, MONTH
(1-4)**453**Has operator applied pesticides with organophosphorous content since last field visit? YES ☐ NO ☐

If YES, enter latest application date _____ and name of pesticide _____

Date (_____)

370

Starting Time (Military Time) ...

371

UNIT LOCATION **P. 21—41**

Number of rows along edge of field

Number of paces into field

| UNIT 1 | UNIT 2 |
|--------|--------|
| | |
| | |

1. Is this the same unit that was laid out last month? Check NO if this is the first visit to lay out the unit or if unit is relocated. For unit(s) checked: YES—skip to Item 3. NO—complete Item 2.

Enter — 1 for YES
2 for NO

| UNIT 1 | UNIT 2 |
|--------|--------|
| 301 | 307 |

For broadcast fields, do not complete question 2. Write a comment at the bottom of this page and skip to Item 3.

ROW SPACE MEASUREMENTS **P. 43—45**

2. a. Measure distance from stalks in Row 1 to stalks in Row 2 Feet and Tenths

b. Measure distance from stalks in Row 1 to stalks in Row 5 Feet and Tenths

| | |
|-----|-----|
| 303 | 304 |
| . | . |
| 305 | 306 |
| . | . |

COUNTS WITHIN 15-FOOT UNITS (INCLUDE CLIP AREAS A AND B)**P. 45—48**

3. a. Number of stalks (Total)

b. Number of stalks in prebud stage

c. Number of stalks with buds, flowers, or heads
(Note: 3a must equal 3b + 3c).

| UNIT 1 | | UNIT 2 | |
|--------|-------|--------|-------|
| Row 1 | Row 2 | Row 1 | Row 2 |
| 321 | 322 | 323 | 324 |
| 325 | 326 | 327 | 328 |
| 329 | 330 | 331 | 332 |

4. a. Total number of buds, flowers and heads

b. Number of buds

c. Number of heads with open flower or seedfill

d. Number of heads with flower wilt or beyond with seeds

e. Number of heads with flower wilt or beyond without seeds
(Note: 4a should equal 4b + 4c + 4d + 4e)

| UNIT 1 | | UNIT 2 | |
|--------|-------|--------|-------|
| Row 1 | Row 2 | Row 1 | Row 2 |
| 341 | 342 | 343 | 344 |
| 345 | 346 | 347 | 348 |
| 349 | 350 | 351 | 352 |
| 353 | 354 | 355 | 356 |
| 357 | 358 | 359 | 360 |

GENERAL COMMENTS: _____

(Continued)

FORM B—SUNFLOWER YIELD COUNTS (Cont'd)

Counts Within 15 Foot Units (Cont'd)

5. STAGE OF MATURITY (Circle one stage code per unit) **P. 47**

| | Prebud or Earlier | Budding Visible | Open Flower and Seed Fill | Flower Wilting | Mature, Wet | Harvest Mature | Blank |
|---|-------------------|-----------------|---------------------------|---|-------------|----------------|-------|
| UNIT 1 | 300 1 | 300 2 | 300 3 | 300 4 | 300 5 | 300 6 | 300 7 |
| UNIT 2 | 302 1 | 302 2 | 302 3 | 302 4 | 302 5 | 302 6 | 302 7 |
| If lowest code is 1, 2, or 3, skip to bottom of page and enter name and time. | | | | If lowest code is 4, 5, or 6, continue. | | | |

6. Is Harvest planned within 7 days?

- ☐ YES → Complete Item 9 only (skip 7 and 8)
☐ NO → Continue

MEASUREMENTS WITHIN UNIT 2, ROW 1 **P. 48—49**

7. Measure diameter of all heads counted in Question 4d, Unit 2, Row 1. (Box #355)

Do NOT remove head. Record widest and perpendicular measurements to nearest 1/10 centimeter using cloth tapes. If more than 30 heads, use blank space on right.

| | Widest | Perpendicular | | Widest | Perpendicular | | Widest | Perpendicular |
|-----|--------|---------------|-----|--------|---------------|-----|--------|---------------|
| 1. | . | . | 11. | . | . | 21. | . | . |
| 2. | . | . | 12. | . | . | 22. | . | . |
| 3. | . | . | 13. | . | . | 23. | . | . |
| 4. | . | . | 14. | . | . | 24. | . | . |
| 5. | . | . | 15. | . | . | 25. | . | . |
| 6. | . | . | 16. | . | . | 26. | . | . |
| 7. | . | . | 17. | . | . | 27. | . | . |
| 8. | . | . | 18. | . | . | 28. | . | . |
| 9. | . | . | 19. | . | . | 29. | . | . |
| 10. | . | . | 20. | . | . | 30. | . | . |

PRE-HARVEST CLIPPING **P. 49**

8. Clip first 3 heads beyond Row 2 of Unit 1 which are maturity code 4, 5, or 6 and have seeds. Mail these heads to State Office.

| | |
|-----------------------|-----|
| Total Diameter | 308 |
| Total Number of Heads | 309 |

| HEAD | 1 | 2 | 3 |
|---------------|---|---|---|
| Maturity Code | | | |

• Skip to bottom of page and enter name and time.

CLIPPING OF HEADS WITHIN BOTH UNITS **P. 49—50**

9a. Clip all heads inside the unit (with and without seeds) on Row 1 Unit 1 beginning at the 5-foot stake up to the yellow stake ("Clip Area A"). Count and bag all heads Number of Heads

381

b. Clip all heads inside the unit (with and without seeds) on Row 1 of Unit 2 beginning at the yellow stake (12.5 foot mark) up to the ending stake at the 20 ft. mark ("Clip Area B"). Count and bag all heads Number of Heads

382

MAIL ALL heads to the regional lab. Keep Unit 1 and Unit 2 separate.

| | |
|-----------------------------|-----|
| Ending Time (Military Time) | 372 |
| Status Code | 380 |
| Enumerator Number | 390 |
| Supervisor Number | 391 |

Enumerator _____

Did your supervisor assist you in working this Unit? YES ☐
NO ☐

FORM I: SUNFLOWER RESEARCH MEASUREMENTS, 1984
Complete this form during your Sept. 1, 1984 visit.

| |
|-------------------------|
| YEAR, CROP, FORM, MONTH |
| 4512 |

Date (_____)

| |
|-----|
| 101 |
|-----|

Insert anchor pin for 50 ft. steel tape exactly next to the last stalk, outside the unit, before the 5 ft. stake.

Report measurements at: (feet and tenths)

| | | 5 Ft. Stake | 1st Stalk Past 5 Ft. Stake | Last Stalk Before 20 Ft. Stake | 20 Ft. Stake | 1st Stalk Past 20 Ft. Stake |
|--------|-------|----------------|----------------------------------|--------------------------------------|-----------------|-----------------------------------|
| Unit 1 | Row 1 | 102 . | 103 . | 104 . | 105 . | 106 . |
| | Row 2 | 107 . | 108 . | 109 . | 110 . | 111 . |
| Unit 2 | Row 1 | 112 . | 113 . | 114 . | 115 . | 116 . |
| | Row 2 | 117 . | 118 . | 119 . | 120 . | 121 . |

Comments: _____

Enumerator _____

Enumerator Number

| |
|-----|
| 122 |
|-----|

Supervisor Number

| |
|-----|
| 123 |
|-----|

APPENDIX B

The following two illustrations are for the reader who wishes more specifics to the exact test. They illustrate how the discrete research measurements affect the uniform distribution assumption. Thus, a distribution which reflects the interaction between rounding and the variable plant spacing must be developed. The first illustration is more detailed than the second.

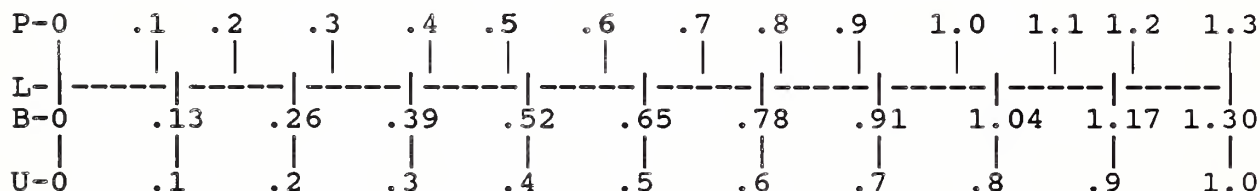
Illustration 1: $I_5=0.7$ feet
 $L_5=1.3$ feet

In order to perform the chi-square test, we divide the unit interval distribution of between-plant spacings into ten subintervals. They are:

$[0,.1), [.1,.2), [.2,.3), \dots, [.7,.8), [.8,.9), [.9,1.0]$.¹

For this observation, the observed stake placement ratio of $I_5/L_5=0.54$ is contained in the sixth subinterval-- $[.5,.6)$.

In order to illustrate the deviation from uniformity caused by rounding, consider the number line (L) below which represents the $L_5=1.3$ foot distance:



The values below the number line (B) represent the boundaries of the ten subintervals of $L_5=1.3$ feet, which can be mapped to the ten subintervals of the unit interval (U). The values above the number line (P) are those 0.1 foot measurements possible in the discrete measurement space with $L_5=1.3$ feet. Notice that each subinterval does not contain an equal number of discrete measurement values. This is due to the rounding process in mapping from the continuous measurement space to the discrete measurement space. For example, the seventh subinterval $[.78,.91)$, which is mapped to $[.6,.7)$, contains the two possible measurements of .8 and .9. These .8 and .9 measurements are rounding results from the intervals $[.75,.85)$ and $[.85,.95)$ respectively. When calculated in conjunction with the other measurements as shown in Table B1, it is evident that there is

¹ In the discrete measurement space, 0 should not be a possible measurement as a consequence of the survey instruction "If a stalk lines up exactly with a stake, it will always be considered past the stake...". However, because of rounding, 0 is a possible measurement.

a 15.4% probability, theoretically, of the 5-foot stake being placed in the seventh (of the ten) subintervals when the adjacent plant distance is 1.3 feet. The probability that each subinterval has of containing a possible 0.1-foot measurement is in the last column of Table B1.

TABLE B1: Theoretical Distribution of Illustration 1,
 $L_5=1.3$ feet, 1984 Sunflower Research

| Ten sub-intervals of $[0,1]$ | Possible discrete meas'ments I_5 | I_5/L_5 | Rounding range of I_5 | Probability ($I_5 \in \text{subint}_i$) |
|------------------------------|------------------------------------|-----------|-------------------------|---|
| $[0,.1)$ | 0,.1 | 0,.08 | $[.0,.15)$ | .115 ² |
| $[\cdot1,\cdot2)$ | .2 | .15 | $[\cdot15,\cdot25)$ | .077 |
| $[\cdot2,\cdot3)$ | .3 | .23 | $[\cdot25,\cdot35)$ | .077 |
| $[\cdot3,\cdot4)$ | .4,.5 | .31,.38 | $[\cdot35,\cdot55)$ | .154 |
| $[\cdot4,\cdot5)$ | .6 | .46 | $[\cdot55,\cdot65)$ | .077 |
| $[\cdot5,\cdot6)$ | .7 | .54 | $[\cdot65,\cdot75)$ | .077 |
| $[\cdot6,\cdot7)$ | .8,.9 | .62,.69 | $[\cdot75,\cdot95)$ | .154 |
| $[\cdot7,\cdot8)$ | 1.0 | .77 | $[\cdot95,1.05)$ | .077 |
| $[\cdot8,\cdot9)$ | 1.1 | .85 | $[1.05,1.15)$ | .077 |
| $[\cdot9,1.0]$ | 1.2,1.3 | .92,1.00 | $[1.15,1.30]$ | .115 |
| $[0,1.0]$ | | | $[0,1.30]$ | 1.000 |

Illustration 2: $I_{20}=0.1$ feet
 $L_{20}=0.7$ feet

For this observation, the observed stake placement ratio of $I_{20}/L_{20}=0.14$ is contained in the second subinterval of the unit interval-- $[\cdot1,\cdot2)$.

Following the same process as detailed in Illustration 1, a number line is provided below for clarification, and the last column of Table B2 displays the probability that each subinterval has of containing a possible 0.1-foot measurement. Notice that two subintervals have no chance of selection.

² Since $I_5=0$ is an endpoint of $L_5=1.3$, the measurement of 0 is rounded from the 'half-size' interval $[0,.05)$, while $I_5=.1$ is rounded from the 'full-size' interval $[\cdot05,\cdot15)$. Therefore, the probability of the first subinterval containing an I_5 measurement is $1.5/13=.115$.

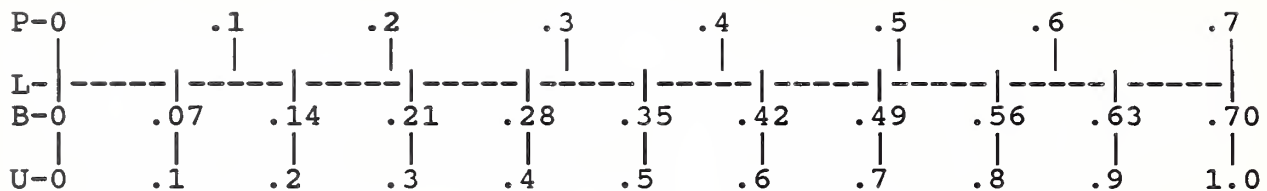


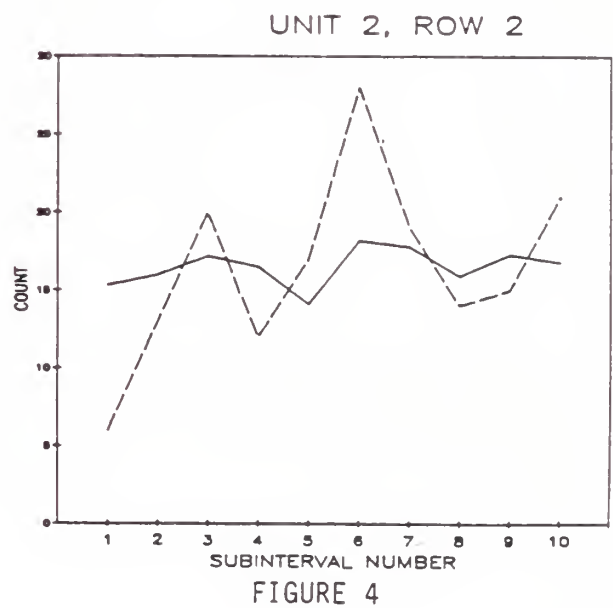
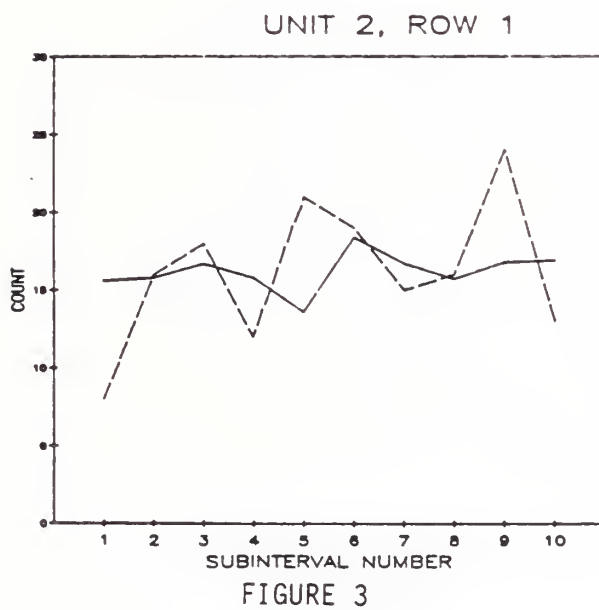
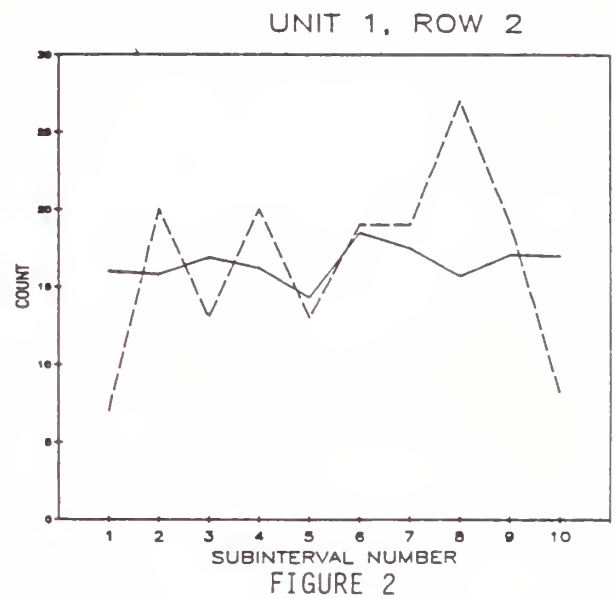
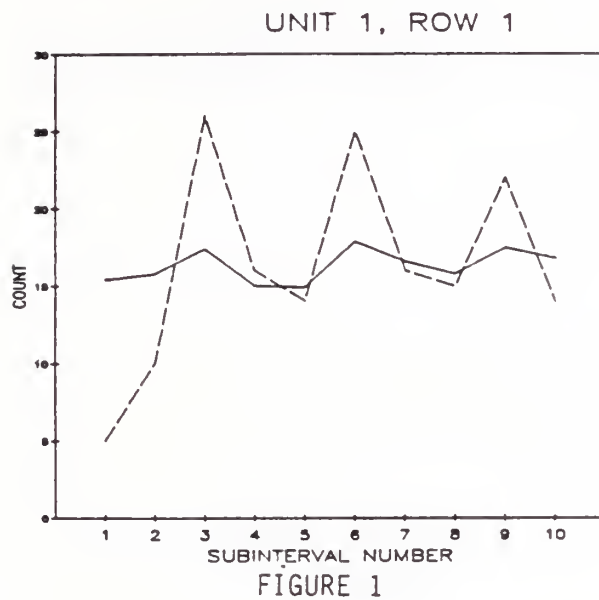
TABLE B2: Theoretical Distribution of Illustration 2,
 $L_{20}=0.7$ feet, 1984 Sunflower Research

| Ten sub-intervals of [0,1] | Possible discrete meas'ments I_{20} | I_{20}/L_{20} | Rounding range of I_{20} | Probability ($I_{20} \in \text{subint. } i$) |
|----------------------------|---------------------------------------|-----------------|----------------------------|--|
| [0,.1) | 0 | 0 | [.0,.05) | .071 |
| [.1,.2) | .1 | .14 | [.05,.15) | .143 |
| [.2,.3) | .2 | .29 | [.15,.25) | .143 |
| [.3,.4) | - | - | - | .000 |
| [.4,.5) | .3 | .43 | [.25,.35) | .143 |
| [.5,.6) | .4 | .57 | [.35,.45) | .143 |
| [.6,.7) | - | - | - | .000 |
| [.7,.8) | .5 | .71 | [.45,.55) | .143 |
| [.8,.9) | .6 | .86 | [.55,.65) | .143 |
| [.9,1.0] | .7 | 1.00 | [.65,.70] | .071 |
| [0,1.0] | | | [0,.70] | 1.000 |

It is now obvious that different subinterval probabilities are associated with varying between-plant spacings. For each observation, the probability of the ratio falling in a particular subinterval of the distance is calculated, and then summed across all observations (for each stake and row of the data set under analysis), arriving at the theoretical distribution. Then, the expected frequencies of the theoretical distribution are compared with the observed frequencies given by the actual stake placement ratio measurements in these same ten subintervals via the chi-square goodness-of-fit test.

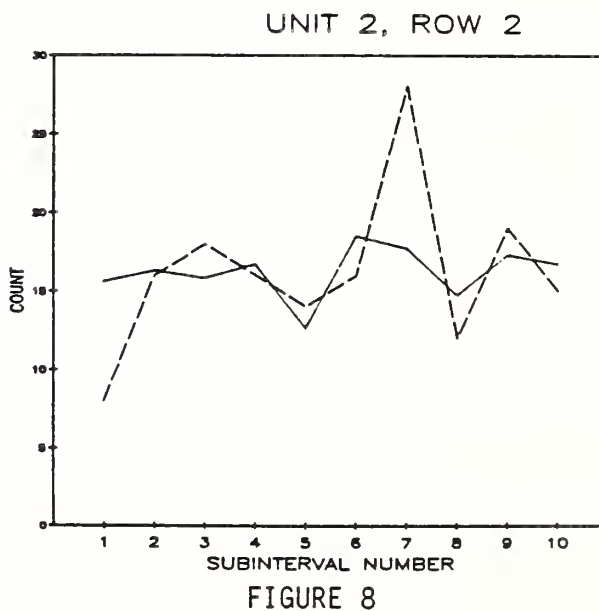
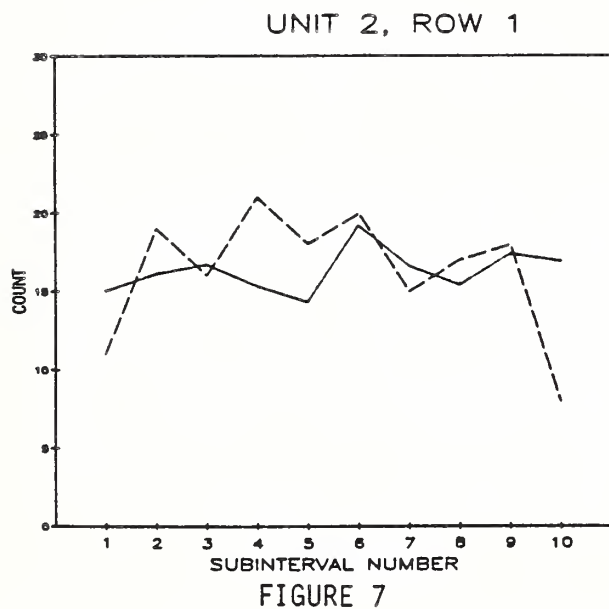
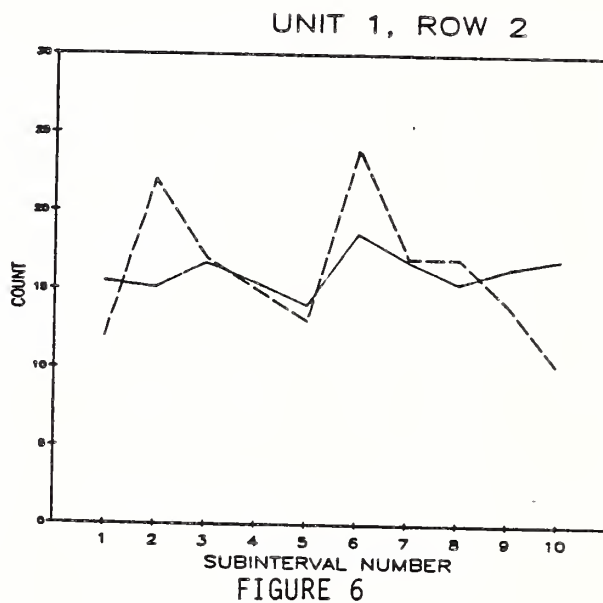
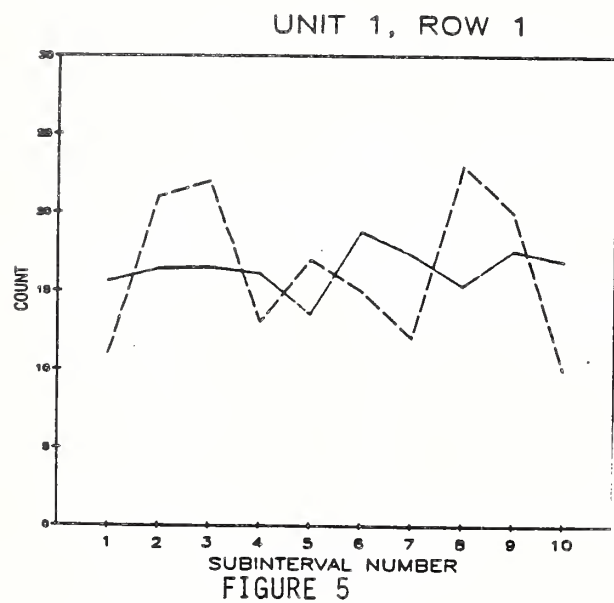
APPENDIX C -- FIGURES 1-4: Comparisons of Expected and Observed 5-Foot Stake Placements, 10 Subintervals, by Unit and Row, 1984 Sunflower Research

(Legend: - Expected -- Observed)



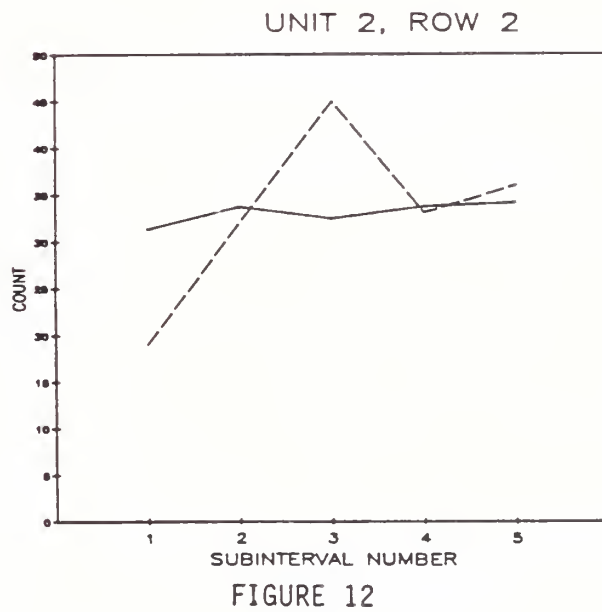
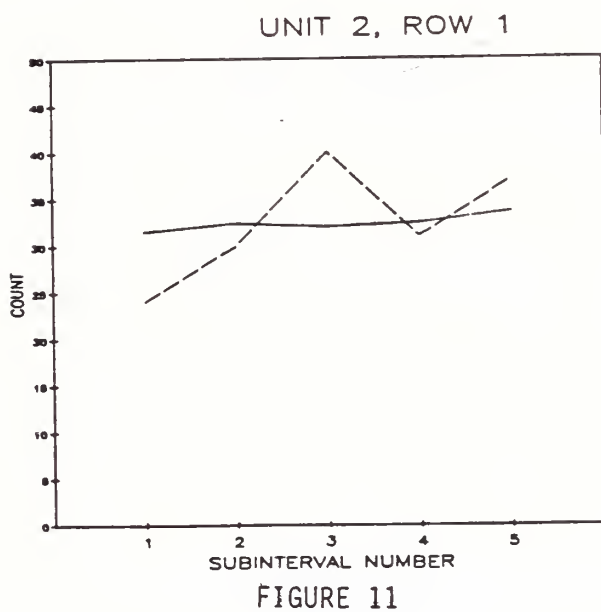
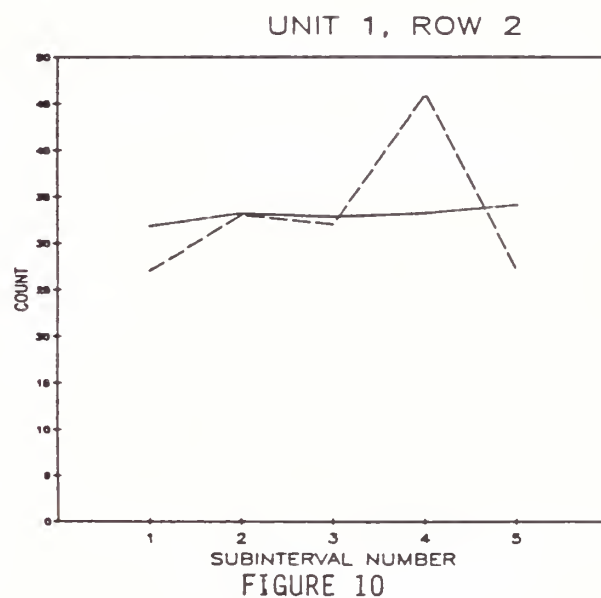
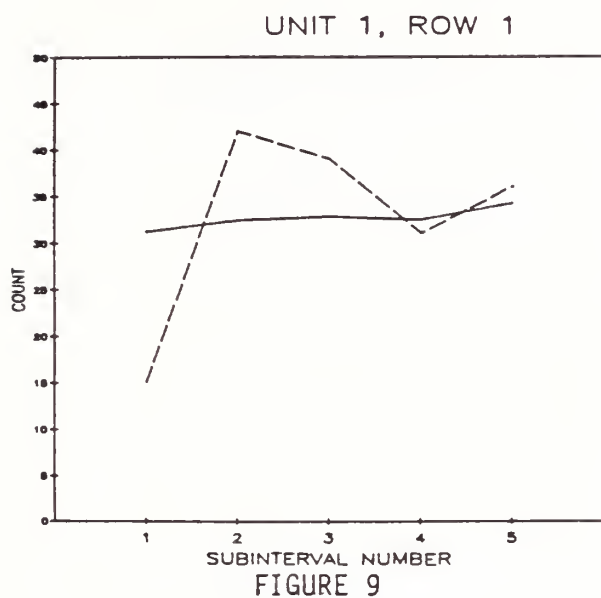
APPENDIX C -- FIGURES 5-8: Comparisons of Expected and Observed 20-Foot Stake Placements, 10 Subintervals, by Unit and Row, 1984 Sunflower Research

(Legend: - Expected -- Observed)



APPENDIX C -- FIGURES 9-12: Comparisons of Expected and Observed 5-Foot Stake Placements, 5 Subintervals, by Unit and Row, 1984 Sunflower Research

(Legend: - Expected -- Observed)



APPENDIX C -- FIGURES 13-16: Comparisons of Expected and Observed 20-Foot Stake Placements, 5 Subintervals, by Unit and Row, 1984 Sunflower Research

(Legend: - Expected -- Observed)

UNIT 1, ROW 1

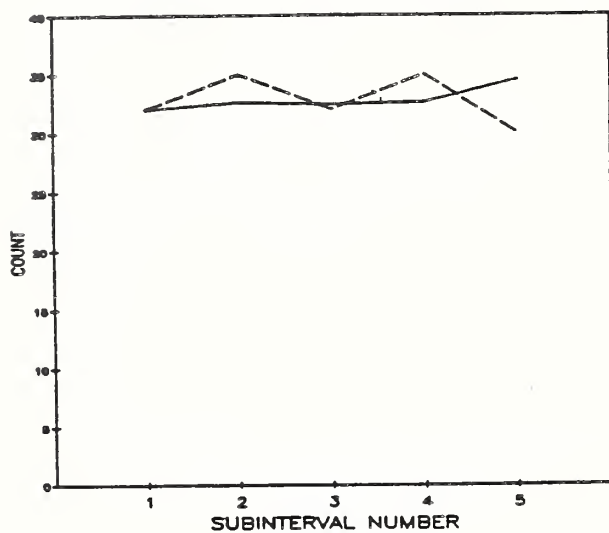


FIGURE 13

UNIT 1, ROW 2

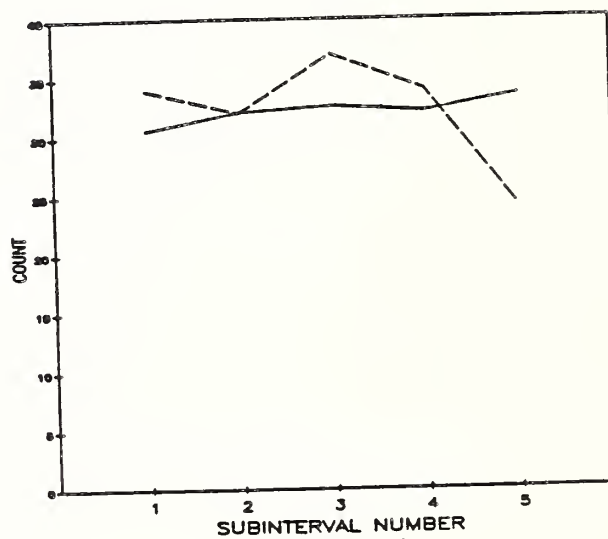


FIGURE 14

UNIT 2, ROW 1

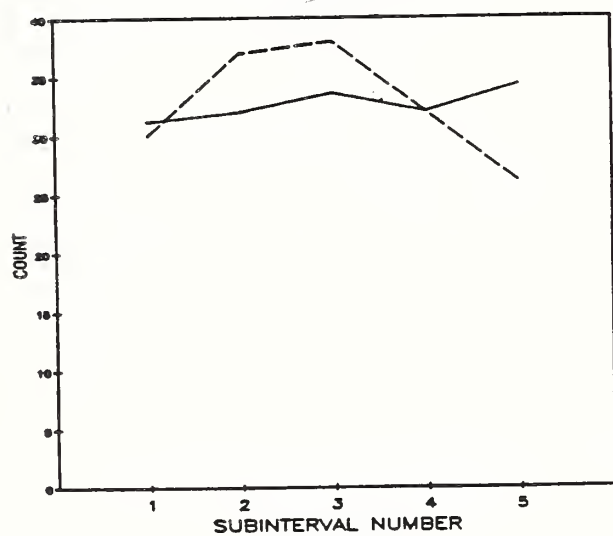


FIGURE 15

UNIT 2, ROW 2

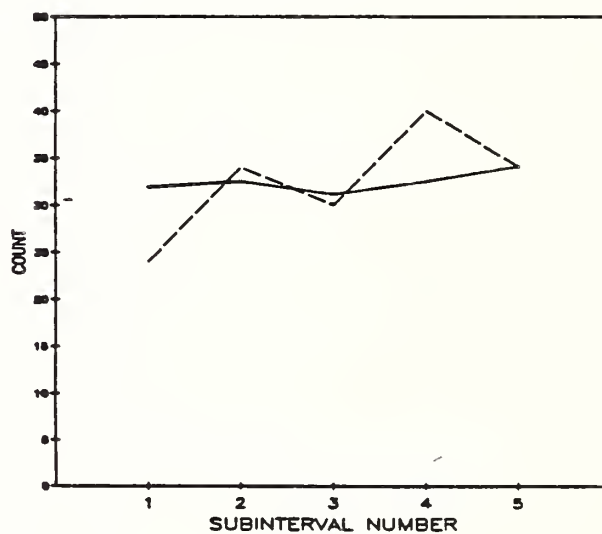


FIGURE 16

